

About Data

Across the Americas, independent scientists measure the flows of carbon between land and atmosphere, using a technique called eddy covariance, and then contribute their data to the AmeriFlux Network. The scientific community uses these measurements to assess responses and feedbacks of terrestrial ecosystems to the environment, including changes in climate, land use, and extreme events such as droughts, storms, or wildfire. AmeriFlux data sets help scientists examine crucial linkages between ecosystem processes and climate responses.

Ecosystem-level field sites acquire continuous measurements from a large number of sensors at high temporal resolution. Data can be collected and processed at different time resolutions, with different units, and quality levels (e.g., filtering criteria and gapfilling). Data can also be derived through process based modeling activities. The data collected at individual sites is sent to an archive operating at a network level. This central network processing involves harmonizing the data, performing uniform data quality checks, and creation high quality and standardized datasets to be distributed to a variety of data users. The data processing includes a number of steps, transforming original measurements from individual sensors to ecologically or micro-meteorologically meaningful quantities. This set of documents describes data formats, data processing levels, metadata, and other relevant information about site and network level data.

Here, a **data variable** is defined as a set of magnitudes or values of a given physically or abstractly meaningful quantity measured by a sensor or derived by processing – e.g., air temperature or quality level. Similarly, a **data variable label** is defined as label assigned to a quantity and used to identify data within a data representation structure such as a file – e.g., *TA*, *TA_F*, *TA_QC*. Data variables can be measured directly by a sensor, or generated by **data processing operations**, which transform one or more data variables into new data variables. Data processing operations can be grouped into stages and data generated at each of these stages define **data processing levels**.

In this context, there are certain aspects that should be easily identifiable when looking at a given data variable. These aspects include:

- **Spatial representativeness:** which area is being represented by a variable (e.g., sensor surroundings or tower footprint);
- **Measured / derived:** whether a data variable is measured directly or derived from measured data through data processing operations or execution of models;
- **Filtered / Unfiltered:** whether a data variable has been filtered for quality purposes or not;
- **Single / multiple sensors:** whether the data variable was originated from a single sensor or a combination of sensors (of either the same type or different types);

- **Process knowledge:** if the processing applied to a data variable includes any ecosystem process understanding that could violate the data independency requirements when used in the context of model validation and parameterization.

The **data variable labels**, the **data processing levels**, and **metadata** associated to a data variable are the main mechanisms to help easily identify these characteristics in a data set and are described next. In more detail, metadata includes information about all aspects listed above. Data variable labels, on the other hand, cover mainly measured/derived, filtered/unfiltered, and single/multiple sensors characteristics. Finally, data processing levels describe spatial representativeness, measured/derived, filtered/unfiltered, and use of process knowledge in the processing.

The following documents details these aspects:

- Data processing levels
- Data variables
- Metadata